How to construct "science commons" together

Shuichi Iwata

1. Introduction

In order to get a solid foundation for constructing "scientific commons", we need to evaluate the feasibility with respect to physical, engineering, economical, social/cultural and political aspects and to make a well-balanced collaborative frameworks for experts to work together. Physical and engineering feasibilities to deal with huge, various kinds of multi-disciplinary data balancing quality of data and data services from data capture to open access by taking advantage of available e-infrastructure have been enlarged quantitatively thanks to evolutions of ICT (Information and Communication Technology). However, such qualitative and semantic issues as data models, standardization, metadata/ontology, qualification of analysis tools and also legal/economic issues like open access, IPR and collaboration schema of different stakeholders have been remaining timeless subjects not so easy to overcome for us all. In this paper, the latter challenging issues are briefly discussed for productive collaborations.

2. Can we set out the guidelines for collaboration?

The process of establishing an inventory of data sets-scientific commons- with an open and inclusive manner for everyone requires us to share an image of wholeness on the final outcome, where and when we need to be flexible and adaptive for the spontaneous evolution of the inventory. We need to work with "neighborhoods" with practical information infrastructures to interact successfully with one another, and to form successful wholes. The necessary guidelines which allow infrastructures for all committed members or stakeholders are to be designed, used and maintained. The infrastructures are expected to give all members "comfortable", "healthy" and "pleasant" space for communication with neighborhoods, and some experts call them "Cloud" which is something beyond traditional information systems and web-based systems.

Many scientists are struggling to manage fresh and ever-growing data of diversity and depth by taking advantage of complex ICT. Databases are as valuable as the quality of the data they store, but there is a problem who pays for the quality. In the nowadays business world, data warehousing are common approaches to improve business information systems, often under labels like CRM, ERP and Supply-chain management, but scientific data are thought in principle as public goods to be shared by everyone. Externalizing data has implications on data access, security, timeliness and availability thus data quality and interoperability may be expected to encompass. There a data governance function is to be established with a recognition of the above important aspects for data services, but the data and information quality require intensive commitments and interactions of data producers and data users coordinated by data service experts.

Into modern architectural concepts such as cloud computing and/or other data services, available data resources are to be re-implemented inheriting valuable contents with a harmonization to the continuously added new data in an ecological and sustainable manner. The deep philosophical semantics and structure in things and processes are to be concerned, which is responsible for the human connection in the world. It requests the origin of the spiritual human dimension to work together with virtual neighborhoods in terms of which gives each person a solid underpinning for step by step actions towards these difficult realms.

3. Our Role

Today people have agreed to enhance cooperation on climate change, clean energy, and the environment overcoming many discrepancies and conflicts in the world, where we need to share high quality data for fruitful discussions and consequent actions. Data activities have become more and more important, and we scientists are expected to contribute to our society through close collaboration to create "science commons" together.

Keywords: database, knowledge infrastructure, CODATA, commons, academic collaboration, data activity
ICSU World Data System

Takashi Watanabe¹*

¹STELAB, Nagoya University

The ICSU World Data System (WDS) has been created through a decision of the 29th General Assembly of the International Council for Science (ICSU) held in October 2008. WDS builds on the 50-year legacy of the ICSU World Data Centre system (WDC) and the ICSU Federation of Astronomical and Geophysical data-analysis Services (FAGS), but WDS will cover wider disciplines ranging from natural sciences to social sciences. Goals of WDS are (from the Constitution of WDS):

* Enable universal and equitable access to scientific data and information
* Ensure long term data stewardship for ICSU programs
* Define agreed data standards
* Establish and oversee the procedures for the review and accreditation of existing and new WDS facilities
* Facilitate better access to data
* Strive for simpler access to data
* Provide quality-assured data and information
* Reduce the digital divide

The WDS Scientific Committee (WDS-SC) has been appointed by ICSU to implement and administer the activities of the WDS. In early 2011, WDS-SC will accept applications from data centers to be members of WDS (http://icsu-wds.org/). The WDS International Program Office (WDS-IPO) will be established in Japan at National Institute of Communications Technology (NICT) in FY 2011. The WDS Conference, which will be the first international meeting of WDS, will be held in Kyoto, on 3-6 September 2011 (http://wdc2.kugi.kyoto-u.ac.jp/wds2011/).

Keywords: ICSU, Data, international collaboration
ICSU World Data System: Hosting International Programme Office in Japan

Yasuhiro Murayama 1*, Ken T. Murata 1, Mamoru Ishii 1, Narihisa Doi 2, Shuichi Iwata 3, Takashi Watanabe 4, Toshihiko Iyemori 5, Hideki Kashioka 1, Toshio Iguchi 1, Hiroshi Kumagai 1

1Natl. Inst. o Info. Communications Tech., 2Keio University, 3University of Tokyo, 4Negoya University, 5Kyoto University

World Data System (WDS) of ICSU (International Council for Science) was decided to be newly established at the ICSU General Assembly in 2008, incorporating legacy of WDS (World Data Center) system and FAGS (Federation of Astronomical and Geophysical data-analysis Services) services. The new system aims at creation of a common globally interoperable distributed data system, or a system of data systems. NICT (National Institute of Information and Communications Technology) of Japan offered hosting International Programme Office (IPO) of WDS in response to ICSU’s announcement to call for a institute hosting WDS-IPO. The ICSU Executive Board decided to accept the offer at its meeting in the end of October 2010. The WDS-IPO will manage and coordinate the establishment and operations of the WDS, and take responsibility for outreach and promotion activities. The IPO will act under the guidance of the ICSU World Data System Scientific Committee. Also NICT is proposing a technical contribution to WDS, incorporating NICT’s potential of network and information system technology, as well as legacy of data archiving activities including NICT’s original atmospheric, ionospheric, and other radio science/space physics databases. In the paper NICT’s perspective will be reviewed to targeting cooperation with Japanese science data community as well as the world wide community of WDS and related bodies.

Keywords: ICSU, WDC, WDS, science information, data
Introduction for the data integration and analysis system (DIAS)

Toshio Koike

1The University of Tokyo

Data Integration and Analysis System (DIAS) was launched in 2006 as part of the Earth Observation and Ocean Exploration System, which is one of five National Key Technologies defined by the 3rd Basic Program for Science and Technology of Japan. The mission of DIAS is:
- to coordinate the cutting-edge information science and technology and the various research fields addressing the earth environment;
- to construct data infrastructure that can integrate earth observation data, numerical model outputs, and socio-economic data effectively;
- to create knowledge enabling us to solve the earth environment problems; and
- to generate socio-economic benefits.

It is expected that there will be a large increase in the volume and diversity of earth observations from inhomogeneous data sources during the next decade. DIAS is developing a core system for data integration and analysis that includes the supporting functions of life cycle data management, data search, information exploration, scientific analysis, and partial data down-loading.

For improving data interoperability, DIAS is developing a system for identifying the relationship between data by using ontology on technical terms and ideas, and geography. DIAS also is acquiring data base information from various sources by developing a cross-sectoral search engine for various data bases.

Keywords: Earth observation data, Data Integration, Information Fusion, Large data storage system, Ontology
The Space Physics Archive Search and Extract (SPASE) Project and the Heliophysics Data Environment

Todd King\(^1\), James R. Thieman\(^2\), D. Aaron Roberts\(^3\)

\(^1\)IGPP/UCLA, \(^2\)Code 690.1, NASA/GSFC, \(^3\)Code 672, NASA/GSFC

The Heliophysics division of NASA has adopted the Space Physics Archive Search and Extract (SPASE) Data Model for use within the Heliophysics Data Environment which is composed of virtual observatories, value-added services, resident and active archives, and other data providers. The size of the data sets and the overall data environment has increased tremendously over the past few years. The SPASE Data Model provides a unifying metadata approach to this complex environment. SPASE has also been adopted by the Canadian Space Science Data Portal (CSSDP), NOAA’s National Geophysics Data Center (NGDC), and recently by Japan’s Inter-university Upper atmosphere Global Observation NETwork (IUGONET). Europe’s HELIO project harvests information from SPASE descriptions of resources. The Planetary Plasma Interactions (PPI) Node of NASA’s Planetary Data System (PDS) is working to map planetary metadata to SPASE for cross-system exchange. All of the data sets in the Heliophysics Data Environment are intended to be described by the SPASE Data Model. Many have already been described in this way. The current version of the SPASE Data Model (2.2.0) may be found on the SPASE web site at http://www.spase-group.org. SPASE data set descriptions are not as difficult to create as it might seem. Help is available in both the documentation and the many tools created to support SPASE description authors. There are now a number of very experienced users who are willing to help as well. The SPASE consortium has advanced to the next step in the odyssey to achieve well-coordinated federation of resource providers by designing and implementing a set of core services to facilitate the exchange of metadata and delivery of data packages. An example is the registry service database shown at http://vmo.igpp.ucla.edu/registry. SPASE also incorporates new technologies that are useful to the overall effort, such as cloud storage. A review of the advances, uses of the SPASE data model, and role of services in a federated environment is presented.

Keywords: SPASE, metadata, heliophysics, interoperability, informatics
The World Meteorological Organization (WMO) is working to organize the WMO Information system (WIS), which is the single coordinated global infrastructure responsible for the telecommunications and data management functions of all programs of WMO. It will be core component of the GEOSS Information System of Systems for weather, water, climate.

According to agreement in the Fifteenth World Meteorological Congress in 2007, WIS is developed in two parallel parts. The Part A is evolution of the Global Telecommunication System (GTS), which has served for time-critical and operation-critical data since 1960s. The Part B is newly-extended information services through flexible data discovery, access, and retrieval to authorized users, as well as flexible timely delivery services.

The WIS network consists of three kinds of centers: the Global Information System Centre (GISC) relays data for global distribution, and centralizes metadata of entire WIS to provide online catalog (clearinghouse); the Data Collection and Product Centre (DCPC) is a hub of regional data distribution and/or provider of specialized product; and all WMO members operates the National Centre (NC). Technical developments for online catalog involves ISO I9115 standard, OAI-PMH and SRU protocols.

This talk will also present lessons from experiences in Japan Meteorological Agency.

Keywords: Meteorology, WMO, Telecommunication, Clearinghouse, Metadata, GEOSS
The state of polar data management; the IPY experience

Masaki Kanao1, Akira Kadokura1, Masaki Okada1, Takashi Yamanouchi1, Natsuo Sato1

1National Institute of Polar Research

The International Polar Year (IPY 2007-2008) was the world’s most diverse international science program. It greatly enhanced the exchange of ideas across nations and scientific disciplines. This sort of interdisciplinary exchange helped us to understand and address grand challenges such as rapid environmental change and its impact on society. The scientific results from IPY only now begin to emerge, but it is clear that deep understanding will require creative use of myriad data from many disciplines. Japan established a national committee for the IPY 2007-2008 in the Science Council of Japan in 2004. A total of 63 projects endorsed by the IPY/IPO (International Program Office) had been planned with Japanese participants. Many of the projects are still under serving as a coordinating platform for post-IPY activities. In the Science Meta-Data Base (SMDB) in the National Institute of Polar Research, Japan (NIPR), a total of 148 metadata sets were accumulated so far with regard to the IPY. Metadata relating to the above IPY endorsed projects, together with other Japanese original and international projects, have been compiled to the IPY Portal in the GCMD (Global Change Master Directory) in NASA (National Aeronautics and Space Administration). In the IPY Portal of GCMD, a total number of metadata descriptions (DIFs: Directory Interchange Format) is more than 90. In the Science Meta-Data Base in the National Institute of Polar Research, Japan (SMDB/NIPR), a total of 148 metadata sets were accumulated so far. The format of metadata is original one, but it includes the items listed in DIFs of AMD (Antarctic Master Directory). There are also links to the corresponding metadata in the AMD for each metadata of the SMDB/NIPR. The SCAR data and information management have worked strongly with the IPY community, and subsequently with the Polar Information Commons (PIC) to help establish the framework for long-term stewardship of polar data and information.

Keywords: International Polar Year, data management, national data center, Antarctic/Arctic Master Directory, SCAR/IASC, Polar Information Commons
International standards and the testbeds for multi dimensional geoinformation

Ryoichi Kouda

1AIST

The technological development of information processing enables the direct treatment of multi dimensional geoinformation as well as two dimensional system to be projected into GIS. World cooperation has been devoted to create the international standards and the testbeds. We review the various disclosures of world geoinformation through the web and supporting international standards with related metadata like as GML and GeoSciML as well as testbeds like One Geology to consider the contribution to the earth and planetary science by the recent trend.

Keywords: Multi dimensional data, geoinformation, international standard, metadata, testbed
Activities for the development of Global Map Version 2

Takayuki Nakamura1∗, Hidehisa Takahashi1, Takeshi Iimura1, Noriko Kishimoto1, Miho Takagi1, Shuhei Kojima1, Masaki Suga1

1GSI of Japan

The Global Mapping Project aims to develop a Global Map through international cooperation of National Mapping Organizations (NMOs) of the world. Global Map is composed of basic geospatial datasets in 1km resolution, covering the whole land area of the globe with consistent specifications. The data consist of eight thematic layers: boundaries, drainage, transportation, population centers, elevation, land use, land cover and vegetation. Global Map Version 1 was released in 2008. The data cover the whole land area of the globe and are used in various fields, namely climate change, disaster, biodiversity, and education, etc.

Global Map data are updated every five years in order to continuously monitor the changes of global environment. Currently Global Map Version 2 data development is underway with a target completion date of 2012. The Version 2 data adopt the data format compliant to ISO (GML3.2.1) and metadata, and several data items and attributes are added to the data. This revision on the specifications was made for the better use of the Global Map data. Serving as the secretariat of the International Steering Committee for Global Mapping (ISCGM), Geospatial Information Authority of Japan (GSI) support the data development by preparing a manual for the data development, metadata editor, and a tool for the data check, among other activities.

At the presentation, the outline of Global Mapping Project and the activities for the development of Version 2 data are reported.

Keywords: Global Mapping Version 2, GML, Global Mapping Project, specifications
Virtual Observatory in Astronomy: Its Construction and Outputs

Masatoshi Ohishi\textsuperscript{1}, Yoshihiko Mizumoto\textsuperscript{1}, Yuji Shirasaki\textsuperscript{1}, Yutaka Komiya\textsuperscript{1}

\textsuperscript{1}NAOJ

Astronomical research utilizes multi-wavelength data that are taken by various ground-based observatories scattered in the world and space-based observatories. Although it has been well known that it is essential to federate such multi-wavelength data, an infrastructure toward realizing the federation was poorly made in the past.

Astronomical Virtual Observatory can be defined as follows: "a collection of integrated astronomical data archives and software tools that utilize computer networks to create an environment in which research can be conducted". Towards a real federation, major astronomical observatories in the world have collaborated in defining standardized protocols and in constructing each virtual observatory. Consequently, as of January 2011, more than 10,000 resources can be accessed from Japanese Virtual Observatory (JVO) that was constructed at the National Astronomical Observatory of Japan (NAOJ). It should be stressed that many astronomical papers were published by using the VOs in the world.

The JVO system has been upgraded continuously. Recently we have implemented the JVOSky system that visualizes observed images and observed data by using the Google API. Further we have experimentally implemented a scalable distributed data analysis system by means of the Hadoop, demonstrating that the net data processing efficiency is some 70 times faster than a single-machine case.

Keywords: Multi-wavelength Astronomy, Database, Information Infrastructure System
A challenge to development of long-term archive for space science data at ISAS/JAXA

Iku Shinohara1*, Keiichi Matsuzaki1, Yukio Yamamoto1, Ken Ebisawa1

1ISAS/JAXA

At ISAS/JAXA, the public service of scientific spacecraft data download via the Internet for space science research community has started since 1998. The ISAS/JAXA’s scientific missions cover extensive fields of space science, e.g. astronomy (X-ray, infrared, radio), solar physics, solar-terrestrial physics, and planetary sciences. At present, datasets from 12 mission projects launched after 1987 archived in the database system, namely, DARTS (Data ARchives and Transmission System). Since major subject of the data service has been the project team use, missions under the operation/development had higher priority of data service implementation. Consequently, the information acquired from the service is sometimes not enough for the general users because the system implicitly suppose the user’s knowledge.

However, in recent years, since some missions unexpectedly continues the operation for a long period or terminated, the perpetuation of datasets or the development of a long-range-data archive becomes a challenge to be solved before the loss of all the important information for the data usage. The collection of the information is also necessary for any users in future. We have begun the activity to achieve the long-term archive of space science data in DARTS, surveying scientific and/or technical backgrounds of each science community in space science. We will present our activity concerning the long-term archive and discuss problems around the data.

Keywords: space science data, data archive
Inter-university Upper atmosphere Global Observation NETwork (IUGONET)

Hiroo Hayashi\(^1\), Yukinobu Koyama\(^2\), Tomoaki Hori\(^3\), Yoshimasa Tanaka\(^4\), Masato Kagitani\(^5\), Atsuki Shinbori\(^1\), Takahisa Kono\(^6\), Daiki Yoshida\(^2\), Shuji Abe\(^6\), Satoru Ueno\(^7\), Naoki Kaneda\(^7\)

\(^1\)RISH, Kyoto Univ., \(^2\)DACGSM, Kyoto Univ., \(^3\)STE Lab., Nagoya Univ., \(^4\)National Institute of Polar Research, \(^5\)PPARC, Tohoku Univ., \(^6\)SERC, Kyushu Univ., \(^7\)Kwasan & Hida Observatories, Kyoto Univ.

The upper atmosphere is considered as a compound system consisting of the mesosphere, thermosphere, ionosphere, and magnetosphere. Although the different atmospheric layers are often referred to as independent regions, they are closely coupled by exchange of materials, momenta, and energies through complicated physical processes. To examine the mechanism of long-term variations in the upper atmosphere, we need to combine various types of ground-based observations made at different locations and altitudes. Each database of such observations, however, has been maintained and made available to the community by each institution that conducted the observations. That is one of the reasons why those data have been used only for studies of specific phenomena. For the same reason some of the observational data have been used by only researcher groups who were involved in the observation campaign and are not easily accessible from the other researchers.

A six-year research project, Inter-university Upper atmosphere Global Observation NETwork (IUGONET), started in 2009 to overcome such problems of data use by the five Japanese research institutes (NIPR, Tohoku Univ., Nagoya Univ., Kyoto Univ., and Kyushu Univ.) that have been leading ground-based observations of the upper atmosphere for decades. We are collaborating to build a database system for the metadata of various kinds of observational data acquired by the global network of radars, magnetometers, optical sensors, helioscopes, etc. The metadata database (MDB) will be of great help to researchers in efficiently finding and obtaining various observational data we have accumulated over many years. The MDB system will significantly facilitate the analyses of a variety of observational data, which we believe will lead to more comprehensive studies of the mechanisms of long-term variations in the upper atmosphere. Moreover, we expect that researchers will become familiar with not only data in their area of expertise but also data from different disciplines by using the MDB. This could promote new interdisciplinary studies of earth and planetary sciences.

The outline of the IUGONET project, along with the current development status and future plan, will be presented.

Keywords: metadata, database, analysis software, upper atmosphere, ground-based observation, interdisciplinary study
Cryosphere Data Archive Partnership (CrDAP)

Hironori Yabuki¹*, Haruko Kawamoto¹, Hideyuki Kitabata¹

¹RIGC, JAMSTEC

The Eurasian cryosphere is an important element of an earth climate system, glacier, frozen ground and snow elements such as large fluctuations in recent years has been focused. IPCC AR-4 Report also describes a number of following and is especially great concern about the social impact. Now in the world snow and ice data are promoted the development by the data center of the United States such as NSIDC (National Snow and Ice Data Center) and NCDC (National Climate Data Center). The actual condition is that frozen ground and snow data does not have an international organization about the data of WMO etc., and present condition grasp and change research do not often become since the international and systematic data archive is very weak. For a better understanding of cold regions of Eurasian cryosphere, it is important to share data over a large area. Eurasia cryosphere, especially in cold regions there are several countries, in order to understand the wide variations in the cryosphere are data management needs of international organizations. The IGOS-Cryosphere and IPY and also has been pointed out the need for it. GEOSS data archiving functions to help improve.

This project is to reveal the reality of global environmental change in Eurasian cryosphere, promoting data collection and catalog information to the public so far has not caught on, clarifying the status of past observations and their data, which aims to make the data public through widespread digitization of data. This project not only in Japan, for the cold regions of Eurasian country, as well as establish a system to promote the release of these data and published research to take over the observed data set, data catalogs and data to researchers widely provide a wake-up.

Keywords: Cryosphere, Database, Metadata
Since 2007 fiscal year Geospatial Information Authority (GSI) of Japan has advanced nationwide development of Fundamental Geospatial Data (FGD) as the main policy of Basic Plan for the Advancement of Utilizing Geospatial Information. The FGD provide positional reference on a digital/electronic map, and consists of 13 items (road edge, building outline and others): GSI's homepage has delivered FGD at free cost since April in 2008. They are produced at the accuracy of map scale 1/2,500 and 1/25,000 within and outside of city planning area, respectively.

Digital Japan Web system served by GSI of Japan takes a browsing role on survey results, and anyone can superimpose geospatial information on the background map data like FGD. Furthermore, the map data are scrollable and scalable. Now GSI promotes usage of the system for national and local government offices and facilitates usage and distribution of geospatial information, base on the principles of Basic Act on the Advancement of Utilizing Geospatial Information.

On the day we will explain these outlines.

Keywords: Fundamental Geospatial Data, Digital Japan, GIS
Recent progresses in observation equipments, analytical techniques and high performance computing technologies have produced huge amounts of geoscience data in various disciplines. However, the data acquisition tools and the data management applications are inherently different among research fields, which eventually produce different data formats even though the observed data might have similar attributes such as longitude, latitude and elevation. Those might be a large barrier to promote cross-disciplinary studies which could give us new insights into the Earth’s dynamics. In this presentation, we introduce examples of eSciences approach in geosciences to handle this problem.

For multidisciplinary data visualization, it is needed to analyze each data format and to acquire a skill to use unfamiliar presentation tools which are not free in general. Here we propose Google Earth as the visualization platform. We have developed tools to help displaying various geoscience data on Google Earth. We have developed software to convert the original data files to a KML file, called ”KML generator”. These generators allow us to visualize various data together on Google Earth without any complicated procedures. We show KML generator for seismic tomography model as an example of our approach. Seismic tomography represents 3-D seismic velocity distribution in the Earth. Lateral heterogeneities of seismic velocity in the mantle are generally assumed to be correlated to temperature anomalies, which can be interpreted as a pattern of mantle convection. Our KML generator visualizes any vertical and horizontal cross sections of the mantle tomographic models, which is useful to understand mantle dynamics.

Our KML generator accepts users to submit their own tomography model at our website to generate KML file for their model. To submit tomographic model, we adopt JSON format, which is proposed as the common data format as tomography model by Federation of Digital Seismograph Network (FDSN). The FDSN is an IASPEI sanctioned organization that brings together the primary operators of broadband seismograph networks throughout the world. The FDSN has successfully acted as an effective organization to coordinate activities in data exchange by introducing Standard for Exchange of Earthquake Data (SEED). The FDSN has proposed data request method based on the email, which enables virtual network data center concept, and considered as a good example of eSciences application in solid earth science.

Keywords: Google Earth, eSciences, KMZ, seismic tomography
Main methodologies of Solar-Terrestrial Physics (STP) so far are theoretical, experimental and observational, and computer simulation approaches. Recently “informatics” is expected as a new (fourth) approach to the STP studies. Informatics is a methodology to analyze large-scale data (observation data and computer simulation data) to obtain new findings using a variety of data processing techniques.

At NICT (National Institute of Information and Communications Technology) we are now developing a new research environment named "OneSpaceNet". The OneSpaceNet is a cloud-computing environment, which connects many researchers with high-speed network (JGN: Japan Gigabit Network). It also provides the researchers rich resources for research studies, such as super-computer, large-scale disk area, licensed applications, database and communication devices. What is amazing is that a user simply prepares a terminal (low-cost PC). After connecting the PC to JGN2plus, the user can make full use of the rich resources via L2 network. Using communication devices, such as video-conference system, streaming and reflector servers, and media-players, the users on the OneSpaceNet can make research communications as if they belong to a same (one) laboratory: they are members of a virtual laboratory.

Keywords: NICT, Science Cloud